DOCUMENT RESUME

ED 137 737 CS 003 336

AUTHOR Katims, Michael; And Others
TITLE The Chicago Mastery Learning Reading Program: An Interim Evaluation.
PUB DATE Apr 77
NOTE 30p.; Paper presented at the Annual Meeting of the American Educational Research Association (New York, New York, April 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Elementary Education; Inner City; *Mastery Learning; Minority Group Children; Program Effectiveness; *Program Evaluation; *Reading Comprehension; *Reading Improvement; Reading Programs; *Reading Research; *Remedial Reading Programs

ABSTRACT

This paper reports the interim results of a quasi-experimental evaluation of the Chicago Mastery Learning Reading Program, an application of Bloom's mastery-learning instructional model to the elementary reading curriculum. Two approximately equivalent classrooms were selected from each of 10 inner-city schools with minority pupils who were below grade level in reading. In some cases, classrooms were randomly assigned to treatment; in the others, teachers and principals made the assignment. The reading comprehension subtest of the Iowa Tests of Basic Skills was used as a pretest and was repeated in January using a different form. Although the control group scored significantly higher on the pretest, a variety of analyses indicated greater gains for the experimental groups. Spring results are expected to confirm the superiority of the experimental program. (Author/AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137737

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE CHICAGO MASTERY LEARNING READING PROGRAM:

AN INTERIM EVALUATION

Michael Katims, Jeffrey K. Smith, Carolyn Steele, John W. Wick

Chicago Board of Education

Paper Presented at the Annual Meeting of the American Educational Research Association, New York, 1977.

CS 003 336

The Chicago Mastery Learning Reading Program (CMLRP) is an experimental program designed to combat the chronic problem of low reading achievement. The goals of the program were to adapt the mastery learning theory and teaching strategies developed by Benjamin S. Bloom and James H. Block[1] to the reading curriculum, and to develop a reading program, based on this theory, which could be easily and cheaply implemented in Chicago's large urban system.

During academic year 1975-1976, a program was developed and simultaneously field-tested in five pilot classrooms.[2] Because the purpose that year was developmental and not evaluative, controlled data were not obtained, and while the results could not be conclusive, they were quite promising. Therefore, an experiment was planned: two classrooms from each of 10 typical inner-city schools were to be selected, with random assignment of classrooms to treatments, and pretest vs. posttest comparisons. The exigencies of real life being what they are, the resulting study is more quasi-experimental than experimental, with the attendant complications for data analysis. Nonetheless, a number of analyses of variance and covariance based on post-scores, gain scores, and estimates of "true" gain all reveal the CMLRP to be superior, though at differing levels of significance.

It is argued that in all of these analyses, the bias inherent in the analysis works against the experimental group since these subjects scored lower on the pretest. Thus, the positive results

[1]B.S. Bloom, "Recent Developments in Mastery Learning", Educational Psychologist, Vol. 10, 1973, pp. 53-57; B.S. Bloom, "Time and Learning", American Psychologist, Vol. 29, 1974, pp. 628-688; B.S. Bloom, Human Characteristics and School Learning, New York, N.Y.: McGraw-Hill Book Co., 1976; J.H. Block, "Operating Procedures for Mastery Learning", in J.H. Block, Ed. Mastery Learning: Theory and Practice, New York, N.Y.: Holt, Rhinehart, and Winston, 1973.

[2]This work was reported at the AERA Annual Meeting in 1976 (Smith and Wick, 1976).

obtained are even more impressive than the probability levels reported with each analysis. It is further argued that if a policy decision were required on the basis of these data, the task would be difficult requiring not only decisions about the level of statistical significance, but also substantive assumptions about the appropriate level of analysis, type of analysis, structure of measurement error, and even about which data to include. It is concluded that while the CMLRP appears to be doing significantly better than the control classrooms, it is fortunate that further data will be available and further analysis possible before any decision-making is required. If the observed effects are real treatment effects, it appears that the Spring 1977 data will definitely confirm the superiority of the CMLRP, independent of the methodological problems.

I. Mastery Learning and Reading

"Mastery Learning" is a phrase of the times, used by many to mean a variety of different things. In its most inclusive meaning, the term could apply to any instructional situation where student achievement is assessed relative to some predetermined criteria as opposed to being assessed relative to other students' performance. Here, however, we are concerned with what Mueller (1976) calls the "mastery learning instructional model" consisting of those organizational and teaching strategies delineated by Bloom and Block. The major components of this model, as summarized by Mueller, are:

1) Formal specification of a comprehensive set of cognitive objectives.

2) Instruction.

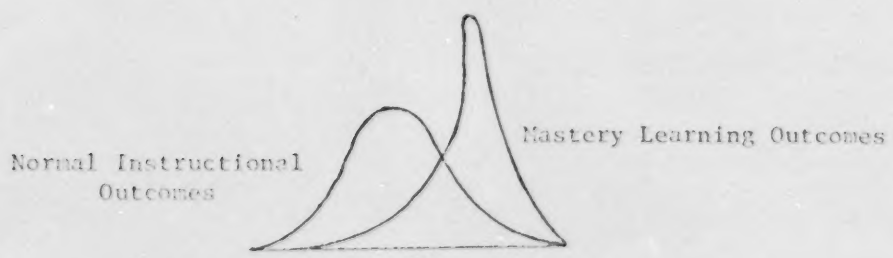
3) Frequent formative/diagnostic evaluation.

4) Corrective or remedial instructional treatments.

5) Criterion-referenced summative evaluation.

The aspects of the model which most clearly differentiate mastery learning instruction from traditional instruction are the regular and systematic use of formative evaluation and corrective instruction. Under this model, students are tested for mastery of the objectives in each instructional unit. Individual deficiencies are identified by this procedure and students who have failed to reach a predetermined level of mastery are given the additional opportunity and assistance to master those objectives. In this way, almost all students have mastered the objectives of the unit and are prepared for the next unit. In addition, because more students have had some success in the course, more students are better prepared affectively (i.e., more receptive, confident) as well as cognitively. Over a series of small experimental mastery learning studies conducted by his students, Bloom (1976) reports these findings: 1) 80-90% of mastery learning students demonstrating mastery on a summative test as opposed to 20% of the control students; 2) greatly reduced correlations between aptitude pretests and summative results under mastery conditions, i.e., final scores less dependent on prior ability; and 3) reduced variance of summative scores under mastery conditions (See Figure A).

FIGURE A



All of these are admirable goals for a program which is designed to help all children achieve an adequate level of learning, independent of any prior characteristics.

However, there is a problem in generalizing about the mastery learning model on the basis of the studies reported by Bloom. Almost all of the studies have used one of three types of content area: mathematics (arithmetic, algebra, matrix algebra, probability), science (biology, psychology, physics), or second language (French, Spanish). All of these are courses which meet three conditions which maximize the effects of the mastery learning model:

1) They are content areas in which beginning students can be assumed to have equal initial ability, if differing aptitudes; thus mastery learning instruction used from the beginning of a course can guarantee that almost all students maintain the necessary cognitive prerequisites.

2) They are content areas in which a sufficient set of objectives can be defined and grouped into learning units, which can be

3) Arranged sequentially because hierarchical relations among the objectives define necessary sequences of some subsets of the objectives.

Thus, these are courses where it is worthwhile to spend extra time on earlier units because the knowledge and skills obtained in these units are <u>necessary</u> for later units. These are courses where a model which maximizes the maintenance of cognitive prerequisites would be expected to produce greater achievement by more students, less variation in achievement, and less relationship between final achievement and initial ability. To summarize, these are subject areas in which the path from no knowledge to mastery can be traced in a series of sufficient (if not necessary) steps which almost every child can be helped to master.

The problem in creating a mastery learning reading course comes in the first of the steps outlined by Mueller:- the definition of a comprehensive set of objectives. Of course, every reading curriculum developer, public and private, does develop a set of objectives, and there is a great deal of consistency among these sets. However, the operationalization of these objectives, and the relations among them, are anything but necessary. To comprehend the implications of operationalization and sequencing for the mastery learning model, consider a typical difference between arithmetic and reading objectives. In arithmetic, having identified the concept of "addition", it is possible to define almost all the formats of addition as well as the factors which influence the difficulty of the problem (e.g., single digit numbers, double digit numbers, fractions, decimal points). Two reasons why it is possible to do this are that the behaviors defined by the objectives have some independence from the rest of arithmetic and they approximate the real-life behaviors where the skill will be used. In reading, a fairly analogous concept is "long vowel sounds". However, there are a variety of quite different ways to test for recognition and production of long vowel sounds (e.g., Which word has the same long vowel sound as this word? Fill the blank in the sentence with a word which makes sense and has a long vowel sound.). All of these operationalizations differ from those in arithmetic in that: 1) they require more than just "long vowel sound" ability, i.e., some reading and reasoning ability, and 2) none of them approximates what one wants the child to do with the skill in reading; in fact, if one is successful in teaching the child to read, he will rarely, if ever, use the concept at all.

Similarly, in an arithmetic course, spending extra time on single digit addition will presumably yield time dividends when doing double digit addition, because most students will have mastered that prerequisite skill. In reading, extra time spent in learning long vowel sounds cannot be expected to directly influence the learning of synonyms, three letter blends, etc., since the concepts are virtually unrelated. Therefore, when creating a mastery learning unit with reading objectives such as this, one is less confident in asking students to spend much extra time in remediation on the particular objectives. And if the mastery model proves effective in teaching children to read, the mechanisms of success will probably be different from the simple "maintenance of cognitive prerequisites" used to account for the majority of the difference in achievement between mastery and control classrooms.

Mueller says that "The mastery model systematizes a series of instructional components which taken together constitute high quality of instruction."[3] That has been the goal in the design of the CMLRP: to use the mastery learning framework to maximize elements which constitute quality instruction, rather than to use it to maximize the relationships between the objectives. Thus, rather than conceiving of the unit on long vowel sounds as merely teaching that concept, the unit is viewed more as "written verbal experience", where the child is taught how to solve yet another kind of verbal problem, and he is required to read, write, and think about words. The stability provided by the regularity of the format (teach-practice-test-remediate) allows students to encounter written language in different forms (sentences, paragraphs, stories, poems,

[3]Mueller, 1976, p. 45

etc.) and successfully master a wide variety of language skills (e.g., making syllables, using a dictionary, identifying the main idea). Furthermore, with the extensive practice and testing provided by the model, both the teacher and the student are able to see the student acquire new skills. This gives confidence to both. Thus, the program is based on the premise that the mastery learning model can work in reading, by 1) providing verbal experience to students in a variety of ways, and in proportion to their need, and 2) by allowing both teachers and students to reap the affective benefits of regular success experiences.

II. The Experimental Treatment: The Chicago Mastery Learning Reading Program

In addition to the problems of adapting the mastery learning model to the reading curriculum, the program had a number of practical constraints imposed upon it as well. Since the goal was to develop a program which could be quickly implemented in hundreds of classrooms throughout the city, a number of the "facts of life" of urban education had to be considered:

1) Since most schools have already invested in reading materials with which their teachers are familiar, they won't be interested in a program that requires extra reading money, whether for additional materials, extra personnel, or additional teacher training.

2) Since schools vary in the type and amount of special equipment (tape recorders, overhead projectors) and the number of aides available, the program had to be usable by a single teacher, in a classroom of up to 35 pupils with no special equipment or on-going assistance.

3) Since the program would be chosen by principals, independent of teachers, the program would have to be sufficient for the minimally competent or experienced teacher, though not inhibitive of the exceptional teacher.

4) Since pupils in the inner city vary greatly in background and ability, and pupils of markedly different ages may be at the same reading level, the program had to be usable by a wide variety of pupils.

5) The program had to be usable within the confines of the regular school day, since it would be unreasonable to assume that pupils requiring extra work would do it outside of class.

The resulting program is one which attempts to utilize the mastery learning format to insure high quality instruction within the confines of these practical limitations as well as the ambiguities of the reading curriculum.

A. The Typical Reading Unit Within the CMLRP

In his book, Bloom defines five elements which taken together determine the quality of instruction: cues - which inform the student what he is to learn and what he has to do in order to learn, reinforcement - to keep him on task, participation - so that the student practices what he is learning, feedback - which allows the student and the teacher to monitor the student's progress, and correctives - remedial instruction for those who need it. In the CMLRP, there are two kinds of units: skills units and comprehension units.[4] However, both are taught by the same framework, which attempts to systematically maximize these elements of quality of instruction. Each unit contains:

1) A guide to initial instruction: This is a thorough "script" to the instruction of the skill being taught. It contains cues for the teacher, and through her to the student, about what skill is being taught, why it is important, how the student will be asked to demonstrate mastery, and what steps he will go through to get there. In most units, there are a series of two or three teaching activities in a sequence designed to lead the child to acquire the skill. Each of these teaching activities has its own guide to instruction.

[4]The objectives on which these units are based are taken from the Curriculum Guide in Reading for the Elementary School. This guide represents Chicago's comprehensive set of reading objectives, divided into four strands (Word Attack, Study Skills, Comprehension, and Literature) of objectives arranged in a longitudinal sequence, and tracing reading development from beginning reading through the end of elementary school. The materials being tested here are based on the end of 2nd grade/beginning of 3rd grade level objectives.

2) Student Activity Sheets: The teacher is provided with ditto masters to create student worksheets that accompany each teaching/learning activity. In doing these individually, and going over them as a class, the student receives additional cues, but more importantly he gets an opportunity to participate, or practice the skill being taught.

3) Formative Test: Each unit contains a formative test for which the correct answers and the mastery levels are provided. This test provides feedback to the teacher and the student about the student's progress in the unit.

4) Remediation: The teacher is also provided with ditto masters for a remedial exercise for those who need it. These are designed to provide additional instruction (usually in a more pictorial or concrete mode), additional practice, and something similar to additional testing. They are designed to be done by the student alone, with the teacher roving through the classroom to provide help or feedback to students at periodic intervals in the exercise. In this way, she can provide help to those who need it, at the point where they are having difficulty, and can also attend to many remediations simultaneously.

For those students who are successful on the formative test, and do not require remediation, two types of reading activity are suggested: 1) the teacher and the school's reading specialist are encouraged to make available to these students high-interest, appropriate level, reading materials (books, magazines, etc.), or 2) additional ditto masters are provided for the creation of "Enrichment Activities" (e.g., puzzles, games, things to make or do) which entertain the student and provide him with further reading experience.[5] These students could also be used to help with remediation (peer tutoring).

It can be seen that after a student has gone through the unit, he will have worked through 1-3 worksheets, a 2-3 page formative test, and for some a 2-4 page remediation. Thus, a child in the CMLRP gets much more verbal experience than students in a traditional reading program, with those students who most require it getting still more experience.

[5]Thus, the frequent criticism that under mastery learning conditions fast students must "wait around" for slow students is somewhat muted in a mastery learning reading curriculum.

B. The Integration of the Components of the CMLRP

As the program now operates, it consists of three basic parts: reading from a basal reader, skills units, and comprehension units. A skills unit typically teaches two or three objectives which can be somewhat concretely defined, while the comprehension units each teach one, more broadly defined skill, such as "making comparisons". Teachers are free to arrange these three components during the week to fit their teaching style and the needs of their classroom, but of the time they devote to reading, they are asked to spend approximately one-third on each of these three activities.

In the first attempts to develop the program and in observations of typical reading instruction, we discovered two interesting facts: 1) Stories in basal readers are independent of the skills instruction that takes place in the accompanying units. For example, if the unit is teaching "ea" words, the story is not necessarily filled with these words. Therefore, there is no logical reason why progress from one story to the next should be dependent on mastering skills. 2) If progress through the different units and the stories are tied together, pupils only read from their readers about once every two weeks since they are otherwise working on and remediating skills. Therefore, we asked teachers to think of their reading week in terms of 10 segments (two per day) instead of the usual five. We asked them to utilize three or four of these segments per week for each of the activities in the program. Thus, a hypothetical week of reading instruction might look like this:

FIGURE B

	M	T	W	T	F
1st Segment	S	S	S	S	C
2nd Segment	R	C	C	R	R

S = Skills Unit
C = Comprehension Unit
R = Reading

In this way:

1) Each classroom reads from its basal readers (or some other material chosen by the teacher) three or four times a week.

2) Each classroom can move through each skills unit and each comprehension unit at its own pace, moving to a new one as they finish each unit.

3) Each school that wants to use the program can use the reading materials it already possesses, requiring in addition only a set of manuals for the teacher, and ditto paper for the pupils (approximately $13.00 per class, per year).

C. Teacher Preparation

One of the practical restrictions on the program is that it must be capable of implementation with a minimum of cost and personnel. Accordingly, the teachers taking part in the experiment were given a bare minimum of preparation. Prior to their one in-service, the teachers were given a copy of Block and Anderson's Mastery Learning in Classroom Instruction, as well as a paper about the program (though it is doubtful that more than two teachers read either document). They then attended a 1 1/2 hour meeting at which they were told about the program, given their materials, and were told about how the different components were designed to fit together. It was clear at the meeting that they found the array of materials, recording forms, and instructions both overwhelming and somewhat intimidating.*

After giving the teachers about a week to become familiar with the materials and to begin to establish the program in their classroom, the staff arranged to meet with each teacher during one or two of the teachers' preparation periods. At these times,

*In the future, when an operating manual exists, this indoctrination should go much more easily.

questions were answered, suggestions made, egos bolstered, etc. The only other staff involvement in the school has been to solve a few confusions between the teachers and their supervisors, and to make regular deliveries of new materials to the teachers. If this amount of preparation is sufficient to enable teachers to use the program effectively, it certainly qualifies as minimal.

III. Design of The Evaluation

In a recent Educational Researcher editorial, Richard E. Schutz differentiated between "academic research on schools", and "operational research in schools". He stated that they differed, but claimed that "The differences lie not in methodology or merit, but in context and communication."[6] However, in some cases, the "in school" context of operational research does affect the methodology of the research: in the selection and attrition of subjects, the choice of criterion variables, and particularly in the freedom available to assign subjects to treatment.

A. Subjects

The targets of the CMLRP are the inner-city classrooms with a high proportion of minority and poverty level pupils, who are rapidly falling behind grade level in reading ability. Accordingly, the schools in the experiment represent the most difficult conditions under which the program will have to operate. The schools range from one having 16% poverty level pupils* (slightly lower than the median Chicago percentage) to one having over 71% poverty level pupils (of 469 schools, this is the 9th highest poverty level).

6Richard E. Schutz, "Research In Schools", Educational Researcher, Vol. 5, #5, May, 1976.

*as determined by the Chicago Title I qualifying procedures.

with five of the schools having more than 40% poverty level pupils. In addition, the schools are, on the average, seven months below the national reading average for 3rd grade pupils, and one full year behind by the end of 4th grade. None of the schools are above the national averages, and only two are above the Chicago averages. All of the schools contain minority populations: eight of the schools virtually have only Black pupils, one school is divided between Caucasian and Latino pupils, and another between Black and Asian-American pupils.

Two classrooms were selected from each school: one to receive the treatment and one to serve as a control. These classrooms were selected because they contained individuals of approximately equivalent reading ability, and they were fairly representative of the school in terms of background charactertistics.

B. Criterion Variables

One way in which "operational research in schools" is often more restricted than "academic research on schools", is in the selection of criterion variables. In Chicago, as in many other areas, the "Reading Comprehension" score from the Iowa Tests of Basic Skills (ITBS) is reading ability. When the results of the annual city wide testing are released, it is a major media event which includes printing scores from every school in the paper. Any new reading program must demonstrate its effect on the Grade Equivalent (GE) score produced by this test to be judged effective.

Therefore, Form 6 of the ITBS was used as a pretest, with each pupil taking the level of test appropriate for his ability. This same form will be used as a final posttest in the Spring. Form 5 of the ITBS was used as an interim posttest in January, and the resulting GE's from each test form the basis for the analysis.

C. Assignment to Treatments

When considering the statistical analysis of non-equivalent control group designs, most statistical experts conclude that the best solution to the methodological problems is never to have gotten into the situation. Randomization, they say, is the only solution to the inescapable bias in matching, gain score analysis, and analysis of covariance - the usual procedures used to compensate for the initial inequalities. Yet when doing research in schools, randomization is not that easy to achieve. Principals have great authority in their domain and may greatly restrict the assignment of classrooms to treatment;* teachers have some choice in whether or not they will participate, and the idiosyncracies of the school (e.g., population, organization) also limit the researcher's freedom.

Therefore, though a fully randomized, treatment by school, experimental design was planned, the resulting design is much more quasi-experimental. In three schools, the principal decided which classroom was to receive the treatment; in two schools the teachers decided among themselves who would use the program; in one school, one teacher's class was already involved in a special program so the treatment classroom was chosen by default. Only in the remaining four classrooms was selection done by random assignment.

D. The Control Treatment

The goal of the study was to compare the CMLRP with a representative sampling of the approaches to reading instruction being used throughout the city, so teachers in the control classrooms were asked to teach reading as they normally do. Some schools use special

*When told that we were theoretically supposed to flip a coin to determine which classroom received the treatment, one principal agreed but said that we had to use his coin. As it later developed, this principal had a two-headed coin, which sent all of his lower achieving students to receive the treatment, and the rest to serve as controls, thus introducing a selection bias to complicate the analysis.

programs such as Computer Assisted Instruction or "Teaching by Typing", some make systematic use of a basal reader system and accompanying tests, and some mix a variety of materials and systems to create a more "home-grown" reading program. Whatever system was being used in the school, the control teachers were simply asked to use that system in the way they normally would to obtain maximum achievement. The only real interference with the instruction in these classrooms was the time required for testing in September and January.

IV. Analysis

In summarizing some current literature about the statistical comparison of non-equivalent control group designs, Cook and Reichardt (1976) make the following recommendation:

> We encourage the analyst to think of himself as an open-minded detective who sifts through considerable messy evidence in numerous analyses and is prepared to draw honestly tentative conclusions in the normal case where one or more alternative interpretations cannot be confidently ruled out.

Essentially, that is the approach taken in this analysis: the data will be described; the various types of analysis will be discussed and reported; and the conclusions of the different types of analysis will be compared.

A. The Data

The data consist of responses from and information about 474 subjects from 10 CMLRP and 9 control classrooms.* The different analyses use different subsets of this population depending on which information is being used, because complete data are not available on all subjects. In addition, data from another control

*One control teacher didn't test until a month after the test date, and these data were therefore considered suspect.

classroom are of dubious value because it appears that membership in this classroom was determined by the pretest*, thus rendering the comparison between this room and the experimental classroom from that school more susceptible to statistical regression (Campbell and Stanley, 1963). Summary statistics and correlations are reported in Appendix A. The most relevant of this information is the higher performance on the pretest by the control group.

B. The Level of Analysis

An argument can be made for either the individual or the classroom as the experimental unit. In one sense, the classroom is the appropriate unit in that the teacher instructs the whole class, and it is the class which moves through the curriculum. On the other hand, in the CMLRP, each student has two paths in each unit (through remediation or not), so that over a large number of units, there are a number of individual "courses" depending on the students' needs and abilities. Statistically, there are advantages to each: using the individual as the unit increases the degrees of freedom and the precision of the analysis, but using the class as the unit eliminates a great deal of error variance. Analyses have been done at both levels.

C. Analysis of Covariance vs. Gain Score Analysis

These are the two statistical models usually used to neutralize initial inequalities between comparison groups. Lord (1960, 1967) points out that these two analyses can produce very different results with the same data because they impose different models

*The principal with the two-headed coin.

on the data. The problem in analysis of covariance is that groups differing at the pretest should be expected to differ at the post-test, and that measurement error reduces the regression coefficients so that the procedure undercorrects for the bias. Thus, the more likely the test is to contain measurement error, the more biased is the analysis of covariance.

In gain score analysis, it is assumed that the differences between groups remain constant. The problem is that subjects who have gained more in the past should be expected to gain more in any given period (i.e., they have a higher "rate of learning"). In this situation, however, because 1) the initial differences are small relative to the time over which they have accumulated (years in school), and 2) because the growth period between testings is small (15 instructional weeks), the bias in this analysis may be small enough to produce no problem.

In Appendix B are reported a variety of gain score and covariance analyses. Appendix B-1 shows the results of one-way analyses based on the 19 observed classroom means; Appendix B-2 shows two-way analyses based on the student as the experimental unit, and estimating for two missing control classrooms, and Appendix B-3 shows the results of the analyses based on the observed data from the 19 classrooms. Appendix C shows the probability levels associated with the treatment effect for each of these analyses.

D. Observed Gains vs. Estimated "True" Gains

One problem with gain score analysis is that measurement error on both tests can produce observed gains which aren't very believable, e.g., gains and losses of over one grade-equivalent in a three month

time span. A possible solution to this problem is to do some estimation, taking into account the fallibility of the test (Lord, 1956, McNemar, 1958). To do this requires some estimate of the test reliability. Since it was unlikely that the reported reliability of the ITBS (r=.904) was accurate for this population, two estimates were made, using approximate reliabilities based on some local data; these are reported in Appendix D. T-tests are computed to show the relative significance of observed gains, as well as estimates based on reliability estimates of .80 and .75. It can be seen from these comparisons that the more believable the tests are, the more significant the difference.

E. Summary

1. All analyses, despite bias, show the CMLRP to be superior to the control group.
2. In general, the gain score analyses make the size of the difference look more significant.
3. If a one-tailed test is deemed appropriate, the program appears to be doing quite well under all analytic conditions, with the exception of the covariance analyses using January GE as the dependent variable and estimating two control classrooms.*
4. Years in school seems to account for a small but consistent amount of variance, which generally improves the appearance of the treatment effect.
5. In the full model analysis: a) In the Gain Score analysis, school is not a significant factor while the interaction is significant (reminding us that there are good teachers

*Since the question asked of a new program is not: Is it better?, but rather: How much better is it?, it seems that a one-tailed test is appropriate.

and bad teachers), and b) In the covariance analysis, school is a significant factor along with the interaction, but this is probably due to the large differences between the school pretest means.

6. Under the mastery learning condition, gains are less correlated with prior abilities (pretest scores). This is an important feature of the theory, and is consistent with the results of previous studies.

V. Conclusions and Implications

It is easy to see why statisticians so strongly recommend randomization in experimental design. It would be much easier if there were only one correct type of analysis, or if different analyses gave the same answer. Here the problem is to determine if there is a treatment effect which is significant enough to warrant completion and implementation of the program. But by comparing the various analyses, with the differing assumptions inherent in each, it is possible to see that "significant enough" doesn't mean: $p < .05$ vs. $p < .10$, or even: one-tailed vs. two-tailed test. The problem is that the effects one observes are a function of the ways of observing as well as a function of the effects themselves. And in choosing one or another way of viewing, a model is imposed for which there may be no definitive tests.

Fortunately, in the case of the CMLRP, no single answer is required yet. There is still a semester of instruction going on, and there is another testing yet to come. Comparison of the second semester's results should be easier because the experimental group, is now slightly (but not significantly) ahead of the control group having overcome the earlier deficit. And if the observed effects

are real treatment effects, they should be large enough to be quite visible over the course of the year, when the size of the gains being sought will be larger than the standard error of the test (as is not the case now).

As far as mastery learning theory is concerned, these results are very promising. As was explained in an earlier section, reading may be the most difficult subject area to adapt to the mastery learning model. If this attempt is successful, it will lend solid support to Bloom's claim that the model can work with any subject. In addition, if the mastery model works in this subject, under these conditions, it would appear that there is much about the model and its mechanisms which are not as yet understood, and for which further research is required.

Summary Statistics and Correlations: Total Sample and Within Groups[a]

1. Total Sample Statistics

Variance	n	$\bar{X}$	SD
Years in School	474	4.854	1.397
September GE	456	3.036	.773
January GE	475	3.218	.823
Gain Grade Equivalent	431	.178	.618

Correlation Coefficients

	YRSCH	SGE	JGE	Gain
Years in School				
September GE	.262**			
January GE	.197**	.699**		
Gain Grade Equivalent	-.079	-.332**	.431**	
Treatment	.142**	-.084	.024	.088

2. Mastery Learning Statistics

Variance	n	$\bar{X}$	SD
Years in School	259	5.030	1.420
September GE	251	2.977	.802
January GE	271	3.235	.854
Gain Grade Equivalent	243	.226	.603

Correlation Coefficients

	YRSCH	SGE	JGE
Years in School			
September GE	.266**		
January GE	.162*	.742**	
Gain Grade Equivalent	-.154*	-.264**	.431**

3. Control Statistics

Variance	n	$\bar{X}$	SD
Years in School	215	4.637	1.346
September GE	205	3.108	.732
January GE	204	3.195	.782
Gain Grade Equivalent	188	.116	.632

Correlation Coefficients

	YRSCH	SGE	JGE
Years in School			
September GE	.286**		
January GE	.300**	.635**	
Gain Grade Equivalent	-.001	-.413**	.441**

*$p < .01$
**$p < .001$

NOTE: The correlation between pretest and gain for the mastery learning group ($r=-.264$) is significantly smaller than that for the control group ($r=-.413$) at the .05 level of significance (one-tailed).

[a]The authors would like to acknowledge the assistance of Arie vander Ploeg in obtaining information from the data.

One-Way Analysis by Observed Classroom Means

1) One-Way Analysis of VARIANCE: Dependent variable in GE Gain

HYPOTHESIS	df	MEAN SQUARE	F	PROBABILITY
Treatment	1	.0927	2.321	$p < .1662$
Error	8	.0399		

2) One-Way Analysis of COVARIANCE: Dependent variable is GE Gain, Independent Variable is "Years in School"

HYPOTHESIS	df	MEAN SQUARE	F	PROBABILITY
Treatment	1	.1221	3.123	$p < .1162$
Error	7	.0391		

3) One-Way Analysis of COVARIANCE: Dependent variable is January GE, Independent Variables are September GE, and "Years in School"

HYPOTHESIS	df	MEAN SQUARE	F	PROBABILITY
Treatment	1	.0520	1.444	$p < .2750$
Error	6	.0360		

NOTE: The analysis of covariance with January GE regressed on September GE is not really appropriate in this case because the within group regression coefficients are significantly different ($p < .10$). This analysis was included for comparison purposes.

Student-Level, Two-Way Analyses Using 10 Mastery and 8 Control Classrooms, and Estimating for 2 Control Classrooms

1) Two-Way Analysis of VARIANCE (Treatment x School): GE Gain is Dependent Variable

HYPOTHESIS	df	MS	F	P
Treatment	1	1.197	3.275	$p < .071$
School	9	.355	.970	$p < .464$
T x S	7	1.030	2.817	$p < .007$
Error	402			

2) Two-Way Analysis of COVARIANCE (Treatment x School): GE Gain is Dependent Variable "Years in School" is the Independent Variable

HYPOTHESIS	df	MS	F	P
Treatment	1	1.126	3.430	$p < .065$
School	9	.274	.748	$p < .665$
T x S	7	.943	2.570	$p < .013$
Error	402			

3) Two-Way Analysis of COVARIANCE (Treatment x School): January GE is the Dependent Variable, September GE is the Independent Variable

HYPOTHESIS	df	MS	F	P
Treatment	1	.554	1.712	$p < .192$
School	9	.631	1.950	$p < .044$
T x S	7	.408	1.260	$p < .270$
Error	401			

4) Two-Way Analysis of COVARIANCE (Treatment x School): January GE is the Dependent Variable, September GE, and "Years in School" are Dependent Variables

HYPOTHESIS	df	MS	F	P
Treatment	1	.631	1.940	$p < .164$
School	9	.640	1.970	$p < .041$
T x S	7	.411	1.270	$p < .266$
Error	400			

Student-Level, Two-Way Analyses Using Observed Data 10 Mastery Learning, and 9 Control Classrooms

1) Two-Way Analysis of VARIANCE (Treatment x School): GE Gain is Dependent Variable

HYPOTHESIS	df	MS	F	P
Treatment	1	1.767	4.654	$p < .033$
School	9	.419	1.083	$p < .374$
T x S	7	.995	2.572	$p < .013$
Error	426			

2) Two-Way Analysis of COVARIANCE (Treatment x School): GE Gain is the Dependent Variable, "Years in School" is the Independent Variable

HYPOTHESIS	df	MS	F	P
Treatment	1	1.952	5.041	$p < .025$
School	9	.312	.806	$p < .611$
T x S	7	.782	2.020	$p < .051$
Error	425			

3) Two-Way Analysis of COVARIANCE (Treatment x School): January GE is the Dependent Variable, September GE are the Independent Variables

HYPOTHESIS	df	MS	F	P
Treatment	1	.072	.206	$p < .650$
School	9	.486	1.400	$p < .185$
T x S	7	.937	2.702	$p < .010$
Error	425			

4) Two-Way Analysis of COVARIANCE (Treatment x School): January GE is the Dependent Variable, September GE and "Years in School" are the Independent Variables

HYPOTHESIS	df	MS	F	P
Treatment	1	.119	.343	$p < .559$
School	9	.502	1.447	$p < .166$
T x S	7	.834	2.405	$p < .020$
Error	424			

Probability Levels Associated With the Treatment Effect
Effect Estimated by Different Types of Analysis

	Analysis with Classroom Means (One-Way)	Analysis Estimating for 2 Control Classrooms (Two-Way)	Analysis Using All Observed Data (Two-Way)
ANOVA: on GE Gain	.165	.033	.071
ANCOVA: GE Gain (YS)	.116	.025	.065
ANCOVA: JGE (SGE)	---	.650	.192
ANCOVA: JGE (SGE,YS)	.275	.559	.164

Two-tailed Test:

$p < .1$ (36.4%); $p < .05$ (18.2%)

One-tailed Test:

$P < .1$ (72.7%); $p < .05$ (36.4%)

Comparison of Differences in Level of Significance Between t-Tests of Observed Gain Scores, and Two Estimates of True-Gain Scores, Based on Different Estimates of Reliability*

School	Mean Gains in Mastery Learning (ML) and Control (C) Classrooms					
	A Observed		B Estimated w/r_{xx}=.80		C Estimated w/r_{xx}=.75	
	ML	C	ML	C	ML	C
1	.600	.082	.357	.189	.221	.140
2	.296	.295	.316	.292	.248	.217
3	.265	--	.246	--	.167	--
4	.319	.116	.282	.125	.199	.111
5	.241	.353	.272	.280	.206	.188
6	.086	.363	.230	.318	.193	.234
7	.067	-.070	.185	.175	.140	.239
8	.391	-.104	.313	.143	.219	.131
9	.119	.141	.217	.237	.168	.188
10	.123	.065	.215	.215	.164	.179
$\bar{G}$	.251	.112	.263	.219	.193	.180
SD_G	.164	.191	.054	.068	.033	.046
$\bar{G}_1-\bar{G}_2$ =	.139		.044		.013	
Sd =	.081		.028		.023	
t =	1.720		1.570		.570	
Two-tailed Probability	$p < .10$		$p < .20$		ns	

*The estimates are based on Lord's (1956) equations for adjusting scores used to measure growth when the variables are fallible:

$\hat{\gamma}_a = (b_{\gamma x.y})\, X_a + (b_{\gamma y.x}) Y_a - (1 + b_{\gamma x.y})\, \bar{X} + (1 - b_{\gamma y.x})\, \bar{Y}$, where

$\hat{\gamma}$ = estimated true gain; X and Y are the observed pre-and post score, and

$$b_{\gamma x.y} = -1 + \left[\frac{(1 - r_{xx})\ (1 + r_{xy}/f}{1 - r^2_{xy}}\right] \qquad b_{\gamma y.x} = 1 - \left[\frac{(1 - r_{xx})\ (1/f^2 + r_{xy}/f)}{1 - r^2_{xy}}\right]$$

$f = S_y/S_x$

REFERENCES

Block, J.H., "Operating Procedures for Mastery Learning", in J.H. Block, Ed. Mastery Learning: Theory and Practice., New York, N.Y.: Holt, Rhinehart, and Winston, 1973.

Block, J.H. and Anderson, L.W., Mastery Learning in Classroom Instruction, New York, N.Y.: Macmillan Publishing Co., Inc., 1975.

Bloom, B.S., "Recent Developments in Mastery Learning", Educational Psychologist, Vol. 10, 1973, pp. 53-57.

Bloom, B.S., "Time and Learning", American Psychologist, Vol. 29, 1974, pp. 628-688.

Bloom, B.S., Human Characteristics and School Learning, New York, N.Y.: McGraw-Hill Book Co., 1976.

Campbell, D.T. and Stanley, J., Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally, 1966.

Campbell, D.T. and Erlebacher, A., "How Regression Artifacts in Quasi-Experimental Evaluation Can Mistakenly Make Compensatory Education Look Harmful", Compensatory Education: A National Debate, Vol. 3, The Disadvantaged Child, J. Hellmuth (ed.), New York, N.Y.: Brunner Mazel, 1970.

Cook, T.D. and Reichardt, C.S., "Guidelines: Statistical Analysis of Non-Equivalent Control Group Designs: A Guide to The Literature", Evaluation, Vol. 3, #1 and 2, 1976, pp. 136-138.

Curriculum Guide in Reading for the Elementary School, Board of Education, City of Chicago, 1974.

Lord, F.M., "Further Problems in the Measurement of Growth", Educational and Psychological Measurement, XVIII (1958), #3, pp. 437-451.

Lord, F.M., "The Measurement of Growth", Educational and Psychological Measurement, XVI (1956), pp. 421-437

Lord, F.M., "A Paradox In The Interpretation of Group Comparisons", Psychological Bulletin, 1967, V. 68, #5, pp. 304-305.

Lord, F.M., "Statistical Adjustments When Comparing Pre-Existing Groups", Psychological Bulletin, 1969, Vol. 72 #5, pp. 336-337.

Lord, F.M., "Large-Sample Covariance Analysis When the Control Variable Is Fallible", Journal of the American Statistical Association Vol. 55, 1960, pp. 307-321.

McNemar, Quinn, "On Growth Measurement", Educational and Psychological Measurement, Vol. XVIII, #1, 1958, pp. 47-55.

Mueller, D.J., "Mastery Learning: Partly Boon, Partly Boondoggle", Teacher's College Record, Vol. 78, #1, Sept. 1976, pp. 41-52.

Schutz, Richard E., "Research in Schools", Educational Researcher Vol. 5, #5, May, 1976.

Smith, Jeffrey K. and Wick, John W., "Practical Problems of Attempting to Implement A Mastery Learning Program in a Large City School System", paper presented at the Annual Meeting of the AERA, San Francisco, 1976.